**Project Problem and Hypothesis**

* What's the project about? What problem are you solving?

Paid Media Mix Optimization – How much spend to allocate to different marketing channels in order to drive the maximum pay subscriptions.

* Where does this seem to reside as a machine learning problem? Are you predicting some continuous number, or predicting a binary value?

Predict how many continuous pay subscriptions can be acquired based on the optimal spend on different media channels.

* What kind of impact do you think it could have?

Saving marketing money by applying it to the most effective multi-channel model to drive low cost and high volume pay subscriptions.

* What do you think will have the most impact in predicting the value you are interested in solving for?
  + Available marketing budgets
  + User signup data
  + Channel effectiveness (comparing channels against each other)
  + Point of diminishing returns
  + Ad network supply (inventory volume) and demand (advertiser inquiries)

**Datasets**

* Description of data set available, at the field level (see table)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Week | Facebook Spend | Instagram Spend | Snapchat Spend | YouTube Spend | SEM Spend |
| 06/10/17 | $5,000 | $3,000 | $12,000 | $4,000 | $3,000 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Facebook Pay Subscriptions | Instagram Pay Subscriptions | Snapchat Pay Subscriptions | YouTube Pay Subscriptions | SEM Pay Subscriptions | Total Pay Subscriptions |
| 2,600 | 1,468 | 7,800 | 956 | 2,000 | 14,824 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**Domain knowledge**

* What experience do you already have around this area?
  + Multi-channel attribution
  + Campaign bidding optimization
  + Marketing context such as creative fatigue, ad relevance, etc.
* Does it relate or help inform the project in any way?

Yes, understanding the context of the campaigns, and underlying logic of how campaign performances are recorded, can help to interpret the data properly.

* What other research efforts exist?
  + Use a quick Google search to see what approaches others have made, or talk with your colleagues if it is work related about previous attempts at similar problems.

https://analyticsartist.wordpress.com/2014/08/17/marketing-mix-modeling-explained-with-r/

http://www.b-eye-network.com/view/17152

* + This could even just be something like "the marketing team put together a forecast in excel that doesn't do well."

Marketing team currently uses previous week’s pay sub data to run a rough estimation of budget allocation and pay sub forecast on future pay sub numbers in excel. A more accurate logistical model can help optimization of marketing spend.

* + Include a benchmark, how other models have performed, even if you are unsure what the metric means.
  + Have R2 be greater than 0.85
  + All spend metrics are significant (P < 0.05)

**Project Concerns**

* What questions do you have about your project? What are you not sure you quite yet understand? (The more honest you are about this, the easier your instructors can help).
  + Can I
    - 1) use spend data to forecast the # of pay subscriptions?
    - 2) predict the best allocation within each media mix in order to get the optimal result, considering that there will be diminishing returns at some point?
* What are the assumptions and caveats to the problem?
  + What data do you not have access to but wish you had?
  + Have linear multi-touch point data instead of the last touch point data.
  + Other factors such as ad-network demand and supply condition that effects the bidding.
  + What is already implied about the observations in your data set? For example, if your primary data set is twitter data, it may not be representative of the whole sample (say, predicting who would win an election)
  + Spend can’t be unlimited.
  + Data is stale after one year.
* What are the risks to the project?
  + What's the cost of your model being wrong? (What's the benefit of your model being right?)

Poor acquisition result.

* + Is any of the data incorrect? Could it be incorrect?

The potential incorrect part, if any, would be the incorrect record of pay subscriptions, given that it didn’t take into consideration the cancel state.

**Outcomes**

* What do you expect the output to look like?
  + The amount of budget toward each channel, and the pay subscriptions that each channel will bring in.
  + Moving forward, the model has the ability to continue learning from the latest data to update its predication.
* What does your target audience expect the output to look like?
  + Recommendation amount of weekly marketing budget toward each channel.
  + Predict pay subscriptions from each channel.
* What gain do you expect from your most important feature on its own?

Spend amount.

* How complicated does your model have to be?

Two logistic regression models

1) predict the # of pay subscriptions, given the budget.

2) predict the allocation of the budget among different channels.

* How successful does your project have to be in order to be considered a "success"?

Have R2 greater than 0.85

* What will you do if the project is a bust (this happens! but it shouldn't here)?

Document what went wrong and continue to iterate even after the class has completed.